



Indicizzare concetti e/o *named entities*

Pino Buizza

Il processo di indicizzazione, nelle sue varie forme, produce rappresentazioni concettuali dei documenti, costituendo sistemi di recupero dell'informazione, cioè di documenti per la loro valenza concettuale. Diversamente, la ricerca per parole o sequenze di caratteri, su testi completi o su registrazioni formalizzate o no, punta sulle espressioni linguistiche attraverso le quali i discorsi o i documenti sono registrati. Ma anche la tradizionale indicizzazione si serve di termini che non designano propriamente concetti: i nomi propri che designano entità individuali. Indagando la natura e il valore dei nomi propri, e delle entità da questi denominate, questo studio vuole coglierne il posto nell'indicizzazione assegnata, se sia assimilabile *tout court* a quello dei termini di concetto o se e quali differenze richiedano anche attenzioni o trattamenti diversificati.

Morfologia

La coppia "concepts or named entities" appare nel principio dell'intestazione uniforme, il primo dei principi dell'IFLA elencati in *Principles underlying subject heading languages (SHLs)* (Lopes e Beall). Il principio si riferisce ad entrambi i termini della coppia nello stesso modo. Nel successivo principio del nome "i nomi di persone, luoghi, famiglie, enti collettivi e opere" sono citati per raccomandare l'uso



della stessa forma del nome usata nei cataloghi per autori e titoli. Non viene indicata nessun'altra differenza fra i due elementi della coppia. Si ritiene quindi, e di fatto si pratica comunemente nella soggettazione, che le *named entities*¹ abbiano un trattamento separato nei sistemi di controllo d'autorità e negli archivi d'autorità. Ci si limita così a considerare quanto riguarda il lessico, la forma delle intestazioni, cioè la morfologia o il lato formale dell'indicizzazione.

In realtà, pure entro questi limiti, la materia non è così semplice, per almeno un paio di motivi. Innanzi tutto, i punti di accesso creati per la catalogazione descrittiva e le entità che essi rappresentano stanno sempre a sé, puntano all'entità in quanto tale, mentre in soggettazione, per la complessità del tema di un'opera, le stesse entità possono essere legate ad altri termini in una stringa di soggetto o incorporate in un'espressione sintagmatica (*phrase*). Sorge, per esempio, il quesito se un nome di persona che è formalizzato con l'inversione debba mantenerla o debba assumere la forma diretta quando non sta nella posizione di parola d'ordine (es., *Letteratura drammatica italiana – Influssi di Shakespeare, William*, o *– Influssi di William Shakespeare*).² D'altro canto, non ogni entità rappresentata da un nome proprio è presente negli archivi d'autorità di nomi. Per le persone e gli enti che non hanno responsabilità di opere, espressioni, etc. (es. *Alessandro Magno*), si possono almeno adottare gli stessi criteri indicati dalle regole di catalogazione per scegliere la forma

¹Non esiste un equivalente italiano corrente di questa espressione: si propone, e di seguito si adotta "entità denominata/e". È incerta anche la sua copertura, che qui si limita alle "entità designate da un nome proprio".

²La forma indicata dal *Nuovo soggettario* è la prima delle due, con la preposizione fra parentesi, a mantenere la separazione fra i due termini, v. l'esempio: *Influssi [di] Ovidius Naso, Publius* (Biblioteca nazionale centrale di Firenze, *Nuovo soggettario: Guida al sistema italiano di indicizzazione per soggetto, Prototipo del Thesaurus* 123). Il caso non si presenta nella tradizione di *Library of Congress Subject Headings (LCSH)* e in *Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU)*, che spezzano l'enunciato di soggetto in stringhe distinte.

del nome. Per gli oggetti, che pure non sono registrati negli archivi d'autorità di nomi (es. *la Venere di Milo*), nelle regole di catalogazione non sono proprio indicati criteri di sorta e il nome deve essere scelto seguendo le fonti di riferimento o altre liste controllate, e secondo i criteri generali per la scelta dei termini preferiti in un vocabolario controllato.

Differenti caratteristiche

A parte questi cenni preliminari, nel presente studio l'interesse è rivolto a un ambito concettuale, più che alla forma dei nomi. Quali sono le differenze fra concetti e entità denominate, per quanto riguarda la teoria e i modelli dell'indicizzazione? se ce ne sono. Tenterò una semplice analisi di concetti e entità denominate per esaminare la loro natura e la correttezza della loro separazione o identificazione.

«Concetto è un'unità di pensiero. Si forma dalla combinazione mentale di alcune o tutte le caratteristiche di un oggetto concreto o astratto, reale o immaginario», o astraendo le caratteristiche comuni di oggetti concreti o astratti, reali o immaginari simili fra loro. «I concetti esistono nella mente come entità astratte indipendenti dai termini usati per esprimerle» e dalla natura concreta o astratta, reale o immaginaria dell'oggetto o degli oggetti che li hanno generati (ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* 4.1).³

Le entità denominate hanno la doppia caratteristica di essere entità individuali (*unique entities*) e di essere identificate da un nome proprio (*named*). Non sono nominate come tali negli standard per i thesauri, dove entità individuali e nomi propri sono trattati

³La definizione è coerente con quella più asciutta dello standard ISO 2788, semplicemente: "un'unità di pensiero", e del BS 8723 e del nuovo ISO 25694-1 che confermano la forma breve e aggiungono in una nota la seconda frase citata sopra.

marginalmente: dal punto di vista dell'inclusione, non necessaria, del controllo della forma, necessario ma non regolamentato, delle relazioni esemplificative (ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* 2.3; 6.3.3; 6.6.7; 8.3.2; ISO/CD 25694-1. *Thesauri and Interoperability with other Vocabularies. Pt. 1: Thesauri for Information Retrieval* 7.6.9; 11.2.4).

Ispirandoci alla classica lezione di Ranganathan (*Prolegomena to Library Classification*), possiamo considerarle secondo tre piani distinti: il piano concettuale (*idea plane*), il piano verbale (*verbal plane*) e il piano notazionale (*notational plane*). Sul primo, ogni entità unica è un'istanza della classe alla quale appartiene; condivide con gli altri membri le specificità della classe, ma non le proprie caratteristiche accidentali, la propria identità. Il concetto è la classe di questi individui ed è definito dalle caratteristiche condivise.

Sul piano verbale, le lingue naturali usano nomi comuni per i concetti, per gli insiemi di individui e per gli individui non identificati (in genere con l'articolo indeterminativo, nelle lingue che dispongono di articoli di diverso tipo), mentre per gli individui identificati usano nomi comuni con l'articolo determinativo o con altri determinativi sufficienti nel contesto, o con descrizioni definite, oppure, infine, con nomi propri. I linguaggi controllati di indicizzazione verbale, in funzione dell'efficacia della ricerca, formalizzano in modo uniforme e univoco sia i nomi comuni che i nomi propri, col controllo d'autorità e con relazioni fra termini differenti ma equivalenti o fra nomi propri riferiti alla stessa entità o fra differenti forme dei termini o dei nomi propri. I termini per i concetti sono scelti da liste di intestazioni per soggetto o da thesauri, con il valore semantico di nomi comuni e l'implicita o esplicita dichiarazione delle relazioni gerarchiche e associative fra concetti. I nomi propri degli autori e di altre entità denominate sono scelti seguendo le

regole di catalogazione o le opere di consultazione o altre opportune consuetudini. Non viene stabilito nessun altro legame, oltre ai rinvii fra diverse forme dei nomi e fra diversi nomi della stessa entità individuale.

Il piano notazionale è tipico dei sistemi di classificazione e non si applica nei sistemi verbali, ma per questi, con un'analogia estensiva, si può considerare il piano sintattico, dove la costruzione di stringhe pre-coordinate formalizza rappresentazioni coerenti di soggetti complessi. In proposito, sistemi fondati sul principio di specificità privilegiano l'entità unica come fuoco o elemento chiave di un soggetto complesso e quindi applicano nella sintassi una distinzione fra i concetti, che possono fungere anche da suddivisione, e le entità denominate, che, ad eccezione delle entità geografiche, ne sono di norma precluse. Sistemi fondati su termini e stringhe e sui ruoli logici invece trattano entrambi indifferentemente.⁴

Un'ulteriore differenza di trattamento si riscontra nei formati di registrazione e scambio (sia MARC21 che UNIMARC), dove sono previste codifiche diverse per le entità le cui istanze individuali sono tipicamente designate da nomi propri (nomi di persone, di enti collettivi, di famiglie, titoli con o senza nome d'autore e toponimi nei campi 600-602, 604-605 e 607 di UNIMARC bibliographic). L'impiego di questi codici rispetto al generico "topic" (campo 606) è riservato proprio alle entità denominate, non alla corrispondente categoria di concetti, che comprenderebbe, per es. per le persone, qualsiasi loro qualificazione, come quelle di condizione professionale (es. ingegneri), sociale (es. poveri), etc. Nelle note sul contenuto del campo è riconosciuta la differenza d'uso rispetto agli stessi nomi nei campi delle responsabilità, grazie alla possibilità di aggiungere

⁴La differenza è esemplificata tipicamente dal confronto fra il *Soggettario*, che prevedeva le Suddivisioni per i soggetti biografici, da trattare come voce principale, e il *Nuovo soggettario*, che applica alle persone l'analisi del ruolo logico svolto nell'enunciato di soggetto e ne fissa di conseguenza il posto nella stringa.

termini che specificano il soggetto secondo l'argomento, la forma, il luogo o il tempo.

Entità denominate e nomi propri

In verità la corrispondenza fra entità denominate e nomi propri non è esatta. Ci sono entità uniche rappresentate da un nome proprio non assoluto. Secondo la maggiore o minore percezione della personalità individuale all'interno di una certa categoria di entità, che è la ragione per ricevere un nome proprio, ci sono entità uniche rappresentate o da un nome proprio associato a un nome comune che ne definisce la classe di appartenenza (es., l'uragano Charlie) oppure da un sintagma nominale pure costituito da un nome comune con l'aggiunta di altri elementi distintivi, come le definizioni spaziali e/o temporali (es., il terremoto dell'Aquila del 7 aprile 2009, formalizzato in Terremoti - L'Aquila - 7 aprile 2009).

Al contrario, alcuni nomi propri rappresentano una pluralità, non un'entità unica e individuale, ma sono spesso considerati entità singolari, perché normalmente sono trattati come tali, e non sono consuete opere su una singola istanza. Sono gli articoli prodotti in serie considerati come un singolo modello, che in realtà costituiscono insieme di articoli identici e separati (l'automobile Citroen 2cv, non è lo stesso della mia Citroen 2cv, color beige, targata AB123XZ ...), sono le specie biologiche, cioè le classi di piante o animali con caratteri simili e in grado di riprodursi (*Leontopodium Alpinum*, o Stella alpina, è il nome di tutti e di ciascuno dei fiori di questa specie), e le entità temporali ricorrenti, considerate come una sola entità per le caratteristiche tipiche e permanenti, ma in realtà ripetute molte volte, ad ogni diversa ricorrenza (il Natale come festa rispetto al Natale 2009 o del 2010...).

Specifici tipi di materiali e di prodotti non numerabili possono avere un nome proprio, generalmente accompagnato da un nome comune. Per esempio, "Chianti" o "Bordeaux" sono i nomi propri della regione d'origine dei vini così denominati, per cui questi sono chiamati più propriamente "vino Chianti" e "vino Bordeaux". Diversamente, l'olio d'oliva della stessa regione non è chiamato "olio Chianti", ma "olio del Chianti", o, in linguaggio formalizzato, Olio d'oliva - Chianti. In un sistema di indicizzazione verbale, il dato morfologico della presenza di un nome comune associato al nome proprio garantisce l'inserimento dei termini nella classe di appartenenza (uragano, terremoto, automobile, vino) e attraverso questa nella rete semantica di contesto, mentre i nomi propri che restano assoluti risultano isolati dalla classe e dal contesto (Leontopodium Alpinum, Natale non figurano come fiori o feste).

Linguistica

In ambito linguistico le differenze fra nomi comuni e nomi propri sono state studiate attentamente. È stato indicato un ampio ventaglio di caratteristiche differenti in grammatica e in sintassi, come pure nel parlato corrente e nelle pratiche interlinguistiche (per esempio in traduzione e in traslitterazione). Il comportamento sintattico dei nomi propri corrisponde strettamente a quello dei pronomi personali e dimostrativi, in ragione del valore anaforico. Sorgono difficoltà, come lungo un continuum fra due polarità che presenti una zona intermedia senza una netta linea di confine, nel definire alcune categorie di nomi, se siano da considerare propri o comuni, per esempio i sostantivi monoreferenziali (come la luna, la terra), i nomi di sostanze, valute, mesi e periodi temporali, lingue, malattie, etc. (oro, marco tedesco, febbraio, lo spagnolo, AIDS). Tralasciando i dettagli linguistici, che non ci competono, e le situazioni ibride e

intermedie, che complicano ulteriormente il problema, focalizziamoci su tipi ben definiti e in particolare sugli aspetti semantici, i più interessanti in un discorso sull'organizzazione della conoscenza e sull'indicizzazione per soggetto.

Filosofia del linguaggio

Un dibattito prolungato e appassionato fra filosofi del linguaggio, semiologi, linguisti ha portato a focalizzare due posizioni opposte: della non-significatività e della significatività (o descrittivista).

Da una parte, il nome di un'entità individuale - il suo nome proprio - è considerato privo di significato, è soltanto un referente, un modo per denotare l'entità senza dirne niente. Un nome proprio ha soltanto denotazione e non ha proprietà connotative: «I nomi propri non sono connotativi: denotano gli individui che sono chiamati con quei nomi, ma non indicano, o implicano, nessun attributo come appartenente a quegli individui. Quando diamo a un bambino il nome "Paolo", o ad un cane il nome "Cesare", questi nomi sono semplicemente segni usati per far sì che quegli individui possano diventare soggetti del discorso» sosteneva J. S. Mill (*Sistema di logica deduttiva e induttiva* vol. 1, 88). «Mettiamo un segno, non però sull'oggetto stesso, ma, per così dire, sull'idea dell'oggetto. Un nome proprio non è altro che un segno insignificante che noi, nella nostra mente, connettiamo con l'idea dell'oggetto, cosicché ogni qualvolta il segno incontra il nostro sguardo, o ci viene in mente, possiamo pensare a quell'oggetto individuale» (vol. 1, 91). «Tutti i nomi generali concreti sono connotativi. Per esempio, la parola "uomo" denota Pietro, Giovanna, Giovanni, e un numero indefinito di altri individui dei quali, presi come classe, è il nome. Ma viene applicata a questi individui, perché posseggono, e per significare che posseggono, certi attributi» (vol. 1, 86). Dall'altra parte, ai no-

mi propri sono stati attribuiti vari diversi livelli di significato. Per citarne solo alcuni, Gottlob Frege assegnava anche ai nomi propri un senso (*Sinn*), qualcosa a metà strada fra la denotazione del nome (l'entità stessa) e le rappresentazioni soggettive che possiamo averne («le opinioni circa il senso possono indubbiamente essere differenti fra loro») (Frege, «Senso e denotazione»), comprendendo a volte un significato lessicale, ma assimilabile piuttosto a un significato associativo (pragmatico). Edmund Husserl sosteneva la corrispondenza uno a uno tra estensione (riferimento) e intensione (significato) (Husserl, *Ricerche logiche* II,1), senza peraltro chiarire la natura di questa intensione propria, né come la possiamo definire. Bertrand Russell considerava i nomi propri come descrizioni troncate o abbreviate (Russell, *Introduzione alla filosofia matematica* cap. 16). Per Ludwig Wittgenstein, nel *Tractatus logico-philosophicus*, il nome significa l'oggetto, cioè, i nomi propri semplicemente denotano. Ma più tardi, in *Philosophische Untersuchungen*, il significato di un nome proprio è costituito dalla descrizione che si può fornire della cosa denominata, non dal referente: in questo modo il significato non è fissato. John R. Searle riconosce che i nomi propri non hanno definizioni e le descrizioni si applicano solo a chi porta il nome, non al nome stesso; i nomi propri «funzionano non come descrizioni ma come ganci a cui appendere descrizioni». Essi «sono connessi logicamente con le caratteristiche degli oggetti cui si riferiscono [...] in un certo modo vago» (Searle, «Nomi propri» 256-7). Il legame con un nome comune non costituisce una descrizione definita ma semplicemente una descrizione o predicato (l'Everest è una montagna), che fornisce peraltro un criterio, elastico, di identità; ma essenziale nella discussione è considerare la distinzione funzionale, la «nostra necessità di separare il far riferimento dalle funzioni predicative del linguaggio» (*Atti linguistici* 225). Saul A. Kripke ha ripreso la tesi della non significatività: un nome proprio funziona come un

designatore rigido, se è stato assegnato a un referente ed è stato trasmesso ad altri utenti; i nomi propri sono connessi ai propri referenti attraverso una catena causale di riferimenti da un iniziale atto di battesimo, quando il riferimento fu assegnato mostrandolo e/o descrivendolo (Kripke, *Nome e necessità*).

Linguistica semantica

Recentemente, ponendosi da un punto di vista linguistico, Willy Van Langendonck afferma che la domanda corretta è «in che modo si costruiscono e funzionano i significati?», non se i nomi propri abbiano significato o no. «Nei nomi comuni il significato determina in modo decisivo il denotato, nei nomi propri il significato aiuta a recuperare un denotato già dato. Per motivi pratici possiamo chiamare ‘significato convenzionale’ il significato dei nomi comuni, ‘significato associativo’ quello dei nomi propri. Entrambe le classi di parole inoltre rivelano un significato grammaticale» (Langendonck 38). «Le tradizionali nozioni di intensione ed estensione (reinterpretate cognitivamente) sono necessarie per distinguere i nomi propri dai nomi comuni. Nei nomi propri l’estensione determina l’intensione, mentre nei nomi comuni è l’inverso» (56). Le conclusioni provvisorie di Van Langendonck riguardo al significato dei nomi propri sono (84-86):

- i nomi propri non hanno alcun significato lessicale asserito;

tuttavia, se informazioni presupposte si possono chiamare “significato”,

- i nomi propri hanno significato presuppositivo, e precisamente:

- significato presuppositivo categoriale, a livello di convenzione linguistica stabilita (es. uomo, donna);
- significato presuppositivo associativo, a livello di uso linguistico (a partire dalle descrizioni del referente o dalla forma del nome);
- significato presuppositivo emotivo (sulla base di particolari esperienze o nella forma accrescitiva o diminutiva);
- significato presuppositivo grammaticale (secondo definitzza, numero, genere).

Infine, il referente dei nomi propri è un referente mentale, in primo luogo, anche per i referenti del mondo reale.

La definizione di nome proprio proposta da Van Langendonck:

Un nome proprio è un nome che denota un'entità unica al livello della convenzione linguistica stabilita perché sia psicosocialmente saliente entro una data categoria a livello di base [pragmatica]. Il significato di un nome, se ne esiste uno, non determina (o non più) la sua denotazione [semantica]. Un importante riflesso formale di questa caratterizzazione pragmatico-semantica dei nomi propri è la loro capacità di apparire in costruzioni appositive strette, come *Burns il poeta, il cane Fido, il fiume Tamigi, o la città di Londra* [sintassi]. (87)

Sulla base di questa analisi linguistica, che raccoglie e integra le sfide della filosofia del linguaggio, anche nella posizione della significatività (presuppositiva) un nome proprio non è come un termine di concetto. Non ha definizione, un termine di concetto sì; il suo referente è unico, può essere descritto in molti modi diversi, nessuno esaustivo, un concetto è un'astrazione in sé conclusa dalle caratteristiche di una molteplicità di individui; una descrizione definita del referente è una funzione della lingua diversa dalla funzione di riferimento del nome proprio. Il denotato di un nome proprio

è dato prima della sua descrizione e non necessita di definizione per essere determinato. Esso è sempre, essenzialmente, il referente mentale prima che l'oggetto nel mondo reale. Un nome proprio che rappresenta un'entità unica quindi manca delle caratteristiche di un nome comune o di un sintagma nominale, che rappresenta tipicamente un concetto. Alcune di queste considerazioni linguistiche possono apparire poco pertinenti in un discorso sui linguaggi di indicizzazione, perché quelle si riferiscono alle lingue naturali e alla loro libera espressione nella comunicazione, questi formalizzano i termini e il loro uso in una morfologia e una semantica rigide, in una sintassi povera. Nondimeno alcune risultanze tornano utili nello specifico che segue.

Linguaggi di indicizzazione

Dal punto di vista dei linguaggi di indicizzazione, sia i concetti che le entità denominate possono costituire l'aboutness di un'opera e i loro termini preferiti o nomi propri vengono usati nei sistemi di indicizzazione. Con la differenza che i primi, come avviene in un thesaurus, possono essere posti in relazione l'uno con l'altro secondo le relazioni semantiche fra i loro concetti, cioè quelle considerate permanenti, perché sono dedotte dalle loro definizioni e generalmente dal loro significato lessicale, non da associazioni occasionali nello svolgimento di un discorso. I secondi invece hanno solo relazione con la categoria del loro significato presuppositivo, mentre associazioni asserite con altri sostantivi o nomi pertengono alla memoria episodica autobiografica e all'uso linguistico. I primi trovano facilmente un posto e ruolo nell'organizzazione della conoscenza, i secondi sono esposti alle diverse asserzioni su di loro, alle diverse immagini mentali dei loro referenti, per diverse persone, in contesti diversi. Le opere su un'entità denominata, pur

essendo focalizzate sullo stesso referente, sono particolarmente candidate a fornire descrizioni e asserzioni differenti. Nei thesauri e nei sistemi di indicizzazione per soggetto, tipicamente si usa la relazione esemplificativa per collegare un nome proprio a un termine per la categoria di persone, luoghi o cose, etc. di cui è un'istanza, con l'avvertenza che si tratta di una relazione gerarchica che non è né generica né partitiva (*ANSI/NISO Z39.19-2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* 8.3.2; *ISO/CD 25694-1. Thesauri and Interoperability with other Vocabularies. Pt. 1: Thesauri for Information Retrieval* 11.2.4). Ma, è questa una vera relazione semantica (a priori, indipendente dalle circostanze e dai documenti (*ISO 2788. Guidelines for the Establishment and Development of Monolingual Thesauri* Introduction)), o è solo un buon espediente per riunire gli individui e richiamarli poi insieme più facilmente, senza compiere una ricognizione nominativa dei membri di una categoria? Soltanto la presupposizione categoriale di base dovrebbe essere permanente e indipendente dal contesto, ma l'analisi linguistica ha mostrato che c'è un livello appropriato, fissato da convenzioni linguistiche, che, proprio per le persone, non è "uomo/donna", ma un livello più specifico, come la carica ricoperta, la professione, la nazionalità o altro ancora, che può cambiare nel tempo e non essere pertinente in un altro contesto.

Se si stabilisce una relazione esemplificativa su un livello più specifico, come è necessario con le persone, sorge il problema se un nome proprio possa avere più di una relazione esemplificativa, cioè produrre una poligerarchia (es., Carla Bruni istanza delle modelle e/o delle consorti di capi di stato, Ronald Regan, istanza degli attori e/o dei presidenti degli Stati Uniti d'America). In un sistema che adotti poligerarchie per le relazioni generiche, è ovviamente possibile adottare queste anche per le relazioni esemplificative. Ma è stata proposta anche un'altra via che, senza poligerarchie, consente

una maggiore coerenza del sistema di relazioni (GRIS, Gruppo di ricerca sull'indicizzazione per soggetto 3.4.4.3). Dato che, per le persone, le categorie di appartenenza non sono permanenti, ma sono determinate di fatto dall'aboutness delle opere sulle persone stesse, nessuna relazione esemplificativa viene fissata nel vocabolario, ma una relazione sintattica rappresenta l'istanza direttamente nella stringa di soggetto dell'opera. Nell'intestazione di soggetto il nome proprio non viene registrato da solo, ma di seguito a un termine di categoria che lo contestualizza (es. a p. 88: Architetti: Bernini, Gian Lorenzo, mentre per altre opere avremo: Scultori: Bernini, Gian Lorenzo, o semplicemente: Bernini, Gian Lorenzo). Questa soluzione è in contrasto con il tradizionale principio di specificità dell'intestazione di soggetto, e introduce una sorta di stile classificatorio anomalo, ma è un modo per dare a un nome proprio una significatività relativa, che non ha di suo, e per evitare di legare per sempre un nome proprio a una professione, anche quando non è il caso. Questa proposta opzionale del GRIS peraltro non è stata recepita nel *Nuovo soggettario*, che fonda la propria impostazione e gran parte delle scelte proprio sulla Guida GRIS, e il Thesaurus si limita a presentare «solo a titolo esemplificativo, alcuni termini che possono ritenersi nomi propri, come nomi di vini, di formaggi, di razze di cani, di alcuni eventi storici» (Biblioteca nazionale centrale di Firenze, *Nuovo soggettario: Guida al sistema italiano di indicizzazione per soggetto, Prototipo del Thesaurus* App. A, p. 201). Resta quindi il problema dell'inserimento delle entità denominate all'interno della rete di significati di un sistema di indicizzazione, solo parzialmente ridotto dalla presenza di stringhe di soggetto che associano nomi propri e termini comuni, indicando l'aspetto significativo dell'entità denominata considerato nell'opera (es.: Buonarroti, Michelangelo – Pittura) (4.4.2.1.6).

Sembra invece di utilità puramente formale (morfologica) la

distinzione, già richiamata sopra, presente nelle codifiche dei termini nei formati MARC, sebbene una gestione avanzata dei sottocampi consenta interessanti risultati di ricerca. In riferimento all'ultimo esempio: il recupero di tutte le opere aventi soggetto l'opera pittorica di un singolo artista si ottiene tramite il valore "pittura" nel campo UNIMARC 600\$x (suddivisione comune di un soggetto biografico).

Classificazioni

Nei sistemi di classificazione –solo un rapido sguardo– le entità uniche (o classi di uno) di solito non hanno una notazione dedicata, poiché il concetto di unicità è opposto alla pluralità che una classe comporta, e non ci dovrebbe essere notazione per una classe di uno, anche se eccezionalmente può esserci il caso. Per esempio, nella Classificazione Decimale Dewey William Shakespeare è presente con notazione propria, in ragione della garanzia letteraria che richiede la classificazione distinta delle sue opere; oppure vi compaiono, nelle religioni, singole divinità, personaggi e testi sacri. Di norma un'entità unica sta indifferenziata insieme ad altre entità uniche della stessa specie in una classe che le raccoglie. In un sistema gerarchico-enumerativo può essere classificata in classi diverse secondo il discorso entro cui è trattata, all'interno di diverse discipline, come avviene per i concetti comuni. Ha posto in una o in più classi, ma non può essere recuperata individualmente, perché non è notata né denominata in sé. Salvo che, come nella Classificazione Decimale Universal, alla notazione per una classe possa essere aggiunto come estensione verbale un nome proprio, per precisare che lo specifico soggetto in quelle classi è quella persona (es.: 1(091)MONTAIGNE = Storia della filosofia: Montaigne). Qualunque entità individuale risulta così collocata nel contesto semantico corrispondente al tema dell'opera in cui è trattata, che può essere diverso in opere diverse.

Questa dispersione e contestualizzazione delle entità individuali le equipara ai concetti, ma l'assenza di un nome (o notazione) proprio le dissolve e le riduce se mai al livello di concetti troppo specifici per trovare un posto proprio nella classificazione. Tolto l'elemento identificativo dell'individualità è tolto anche il quesito sulla sua significatività, che è ricondotta tutta al significato della classe attribuita (nella CDD Bernini apparirà opportunamente in 720.92 o 730.92 o in altre classi più appropriate per i diversi riferimenti a lui attribuiti nelle opere classificate).

D'altra parte, in ambito bibliografico, secondo l'insegnamento di Ranganathan, una classe di uno non può nemmeno esistere, per il motivo che anche la classe più specifica può sempre essere suddivisa secondo le articolazioni dei discorsi che la riguardano. Se le classi assegnate, come i nomi propri, hanno referenti mentali, non è né difficile né irrealista dividerle senza spezzare l'unità dell'entità, semplicemente scegliendo alcuni aspetti (una sorta di relazione di appartenenza come parte o proprietà) o aggiungendo connessioni con altri concetti che ne circoscrivono l'ambito (relazioni sintattiche). Nello stesso modo, nei linguaggi d'indicizzazione verbale, un nome proprio fa sì riferimento a un'entità denominata, ma è anche il termine per riunire un insieme di argomenti più ristretti e correlati come parti specifiche del soggetto (es. particolari periodi della vita di una persona, settori della sua attività, o suddivisioni di un'area geografica, capitoli di un'opera etc.) e altri soggetti correlati con l'entità denominata (es. le opere create, le teorie elaborate da una persona, gli eventi a cui ha partecipato etc.).

Oltre l'indicizzazione

Volendo allargare lo sguardo oltre l'indicizzazione semantica, potremmo accennare ad alcune implicazioni di un certo interesse.

Se il nome proprio di un'entità denominata (persona, ente collettivo o opera) è usato nei sistemi di indicizzazione con un significato categoriale, costituito dal suo referente mentale, cioè dall'immagine mentale del referente nel mondo reale, il suo valore e la sua funzione sono diversi da quelli che lo stesso nome ha come elemento d'accesso in catalogazione descrittiva, dove svolge la mera funzione denotativa rispetto al referente nel mondo reale. Non solo, se il referente mentale è teoricamente aperto a raccogliere quasi qualsiasi asserzione, significa che ogni referente mentale ritaglia sulla base delle reali asserzioni adottate un profilo particolare dell'entità denominata. Altrettanto diversi sono il valore e la funzione del nome, se nei linguaggi d'indicizzazione si rimedia all'isolamento semantico delle entità denominate con le relazioni esemplificative o con il contesto della classificazione o altro, mentre in catalogazione le stesse entità sussistono senza legami di sorta ad altro da sé.

Diversità non puramente concettuali, che richiedono considerazione nell'impostare i sistemi di organizzazione e recupero dell'informazione. Le attuali possibilità di ricerca indifferenziata per autore e per soggetto, che includono negli opac una versione moderna del catalogo dizionario, portano i vantaggi di visioni globali sulle entità denominate. Ma il vantaggio informativo è reale se è accompagnato dalla chiarezza, diversamente rischia di generare confusione, e l'autonoma coerenza con le proprie funzioni, quindi la distinzione nella creazione di accessi nominali e semantici e nella loro ricerca è la premessa per l'efficacia di entrambe. Ciò vale tanto più se all'indicizzazione assegnata affianchiamo la derivata e la ricerca libera a testo completo, assistita o no da algoritmi di ricerca che possano aumentare la cosiddetta "rilevanza" dei risultati. Nel campo dell'estrazione dell'informazione in particolare, il riconoscimento delle entità denominate (Named Entities Recognition, NER o riconoscimento e classificazione, Named Entities Recognition

and Classification, NERC) è diventato argomento di studio e motivo per creare strumenti informatici per il riconoscimento automatico. Senza addentrarci in questa tematica, bisogna perlomeno rilevare che, cercando nomi propri in testi completi, si dovrebbe considerare come nel linguaggio naturale un nome proprio possa assumere funzioni assai differenti da quella di denotare il suo referente. Può funzionare secondo metafora o metonimia, per antonomasia, anche reggere una funzione antonimica, oltre ad essere esposto alla molteplicità di attribuzioni per omonimia. In questo senso, e dato che il referente può cambiare di volta in volta, questi casi a livello di uso linguistico dovrebbero essere chiamati lemmi propri, invece che nomi propri, come avviene a livello lessicale; come tali escono dalla nostra indagine limitata all'indicizzazione assegnata, ma non possono essere elusi da studi che includano l'indicizzazione derivata e dall'estrazione automatica dell'informazione. Nondimeno accennare alle ambiguità nell'uso di nomi propri/lemmi propri nel linguaggio naturale conferma che i differenti usi della stessa parola si devono trattare in modi diversi.

In conclusione, ciò che permette ai nomi propri di inserirsi in una rete informativa semanticamente organizzata è la loro funzione di referente mentale rispetto alle entità individuali che essi denominano, non come concetti definiti, ma per l'insieme di asserzioni effettivamente attribuite. Una chiara distinzione fra concetti e entità denominate è utile per comprendere il loro posto fra individualità e contestualizzazione e per adottare nomi propri nell'indicizzazione e nei vocabolari controllati, così come per il rapporto tra elementi d'accesso nominali e semantici, per la considerazione dei nomi propri nell'indicizzazione derivata e nel recupero dell'informazione su testo completo.

Riferimenti bibliografici

- ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, Md: NISO, 2005. (Cit. alle pp. 3, 4, 13).
- Biblioteca nazionale centrale di Firenze. *Nuovo soggettario: Guida al sistema italiano di indicizzazione per soggetto, Prototipo del Thesaurus*. Milano: Bibliografica, 2006. (Cit. alle pp. 2, 14).
- . *Soggettario per i cataloghi delle biblioteche italiane*. Firenze: Il Cenacolo, 1956.
- BS 8723:2005. *Structured Vocabularies for Information Retrieval*. London: BSI, 2005.
- Frege, Gottlob. «Senso e denotazione». *La struttura logica del linguaggio*. A cura di Andrea Bonomi. Milano: Bompiani, 1973. (Cit. a p. 9). Trad. di «Uber Sinn und Bedeutung».
- . «Uber Sinn und Bedeutung». *Zeitschrift fur Philosophie und philosophische Kritik* 100. (1892): 25–50.
- Gnoli, Claudio e Fulvio Mazzocchi, cur. *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference, 23-26 February 2010, Rome, Italy*. Wurtzburg: Ergon, 2010.
- GRIS, Gruppo di ricerca sull'indicizzazione per soggetto. *Guida all'indicizzazione per soggetto*. Roma: AIB, 2001. (Cit. a p. 14).
- Hopkinson, Alan, cur. *UNIMARC Manual: Bibliographic Format*. München: Saur, 2008.
- Husserl, Edmund. *Logische Untersuchungen*. Halle: Niemeyer, 1900-1901.
- . *Ricerche logiche*. Milano: Il Saggiatore, 1968. Trad. di *Logische Untersuchungen*. (Cit. a p. 9). Trad. di *Logische Untersuchungen*.
- ISO 2788. *Guidelines for the Establishment and Development of Monolingual Thesauri*. Geneva: ISO, 1986. (Cit. a p. 13).
- ISO/CD 25694-1. *Thesauri and Interoperability with other Vocabularies. Pt. 1: Thesauri for Information Retrieval*. Geneva: ISO, 2008. (Cit. alle pp. 4, 13).
- Kripke, Saul A. «Naming and Necessity». *Semantics of Natural Language*. A cura di Donald Davidson e Gilbert Harman. Dordrecht: Reidel, 1999.
- . *Nome e necessità*. Torino: Bollati Boringhieri, 1999. Trad. di *Naming and Necessity*. (Cit. a p. 10). Trad. di *Naming and Necessity*.
- Langendonck, Willy Van. *Theory and Typology of Proper Names*. Berlin: De Gruyter, 2007. (Cit. alle pp. 10, 11).
- Lopes, Maria Ines e Julianne Beall, cur. *Principles Underlying Subject Heading Languages (SHLs)*. München: Saur, 1999. (Cit. a p. 1).
- Mill, John Stuart. *A System of Logic, Ratiocinative and Inductive*. New York: Harper, 1843.

- . *Sistema di logica deduttiva e induttiva*. Torino: UTET, 1988. Trad. di *A System of Logic, Ratiocinative and Inductive*. (Cit. a p. 8). Trad. di *A System of Logic, Ratiocinative and Inductive*.
- Ranganathan, Shiyali Ramamrita. *Prolegomena to Library Classification*. 3^a ed. Bangalore: Sarada Ranganathan Endowment for Library Science, 1967. (Cit. a p. 4).
- Russell, Bertrand. *Introduction to Mathematical Philosophy*. London: Allen & Unwin, 1919.
- . *Introduzione alla filosofia matematica*. Milano: Longanesi, 1946. Trad. di *Introduction to Mathematical Philosophy*. (Cit. a p. 9). Trad. di *Introduction to Mathematical Philosophy*.
- Searle, John R. *Atti linguistici*. Torino: Boringhieri, 1976. Trad. di *Speech Acts*. (Cit. a p. 9). Trad. di *Speech Acts*.
- . «Nomi propri». *La struttura logica del linguaggio*. A cura di Andrea Bonomi. Milano: Bompiani, 1973. (Cit. a p. 9). Trad. di «Proper Names».
- . «Proper Names». *Mind* 67. (1958): 166–173.
- . *Speech Acts*. Cambridge: CUP, 1969.
- Wittgenstein, Ludwig. *Logisch-Philosophische Abhandlung*. London: K. Paul, 1922.
- . *Philosophische Untersuchungen*. Oxford: Blackwell, 1953.
- . *Ricerche filosofiche*. Torino: Einaudi, 1967. Trad. di *Philosophische Untersuchungen*. Trad. di *Philosophische Untersuchungen*.
- . *Tractatus Logico-philosophicus*. Torino: Einaudi. Trad. di *Logisch-Philosophische Abhandlung*. Trad. di *Logisch-Philosophische Abhandlung*.

PINO BUIZZA, Biblioteca Queriniana, Brescia
email gbuizza@comune.brescia.it

Buizza, P. "Indicizzare concetti e/o *named entities*". *JLIS.it* Vol. 2, n. 2 (Dicembre/December 2011): 4707-1-4707-21. DOI: [10.4403/jlis.it-4707](https://doi.org/10.4403/jlis.it-4707). Web.

ABSTRACT: A partire da un punto di vista semantico più che morfologico, l'articolo è focalizzato il problema del significato dei nomi propri, con contributi della filosofia del linguaggio e della linguistica semantica. Sono indagate le entità individuali: il loro isolamento all'interno della rete di soggetti e la relazione esemplificativa, il trattamento nelle classificazioni. Le profonde diversità rilevate fra concetti e entità denominate suggeriscono di dichiararle esplicitamente da un punto di vista teorico e di adottare dispositivi che diano risultati unitari ma chiaramente distinguibili nei sistemi di recupero dell'informazione.

KEYWORDS: Concetti; Filosofia del linguaggio; Indicizzazione semantica; Linguistica semantica; Nomi propri.

ACKNOWLEDGMENT: Questo contributo è stato presentato col titolo *Indexing concepts and/or named entities* all'11° ISKO Conference, *Paradigms and conceptual systems in knowledge organization*, Roma, 23-26 febbraio 2010, non pubblicato negli atti, e qui leggermente ampliato.

Submission: 2011-04-13 / Accettazione: 2011-04-16 / Pubblicazione: 2011-06-15.

